

Analisis Diskriminan Kernel dengan Metode *Cross Validation* (Studi Kasus : Faktor-Faktor yang Berhubungan dengan Kejadian Hipertensi pada Puskesmas Usuku Wakatobi Sulawesi Tenggara Tahun 2013)

Rasmi¹, Saleh², La Podje³

ABSTRAK

Pada umumnya metode yang sering digunakan untuk masalah klasifikasi adalah metode Diskriminan Linier. Pada penerapannya, metode diskriminan linier sering melibatkan variabel-variabel yang tidak mengikuti pola distribusi normal, sehingga diperoleh hasil klasifikasi diskriminan linier yang tidak optimal. Untuk mengatasi hal tersebut maka digunakan analisis diskriminan kernel yang menggunakan pendekatan fungsi kernel. Penelitian ini bertujuan mengklasifikasikan pasien yang hipertensi dan yang tidak hipertensi dengan Analisis Diskriminan Kernel yang menggunakan metode *Cross Validation*. Hipertensi sering disebut sebagai penyakit darah tinggi. Melihat angka kunjungan hipertensi dengan jumlah yang cukup besar mendorong untuk melakukan untuk mengklasifikasikan kelompok pasien yang hipertensi dan kelompok pasien tidak hipertensi. Metode yang digunakan dalam pengklasifikasian ini adalah metode *cross validation* dengan membagi data menjadi dua yaitu data training dan data testing dan metode kernel dengan menggunakan fungsi kernel, menentukan *bandwith* optimal dan nilai probabilitas posterior. Berdasarkan nilai probabilitas posterior terbesar akan mengklasifikasikan pasien yang hipertensi dan yang tidak hipertensi. Berdasarkan hasil penelitian yang telah dilakukan diperoleh hasil klasifikasi adalah 68 untuk pasien yang hipertensi dan 32 yang tidak hipertensi. Tampak bahwa nilai *Press's Q* sebesar 84, artinya hasil klasifikasi konsisten.

Kata Kunci: Diskriminan Kernel, Metode *Cross Validation*, Hipertensi.

1. Pendahuluan

Analisis statistika memiliki banyak metode yang dapat digunakan untuk klasifikasi, antara lain metode parametrik. Metode parametrik membutuhkan asumsi-asumsi yang harus terpenuhi sebelum data dapat dianalisis seperti pada metode diskriminan.

Analisis diskriminan bertujuan untuk memisahkan dan mengalokasikan objek pengamatan ke dalam kelompok sehingga setiap objek menjadi anggota dari salah satu kelompok dan tidak ada objek yang menjadi anggota lebih dari satu kelompok. Prosedur analisis diskriminan dilakukan dengan membentuk fungsi diskriminan berdasarkan kombinasi linear dari variabel prediktor yang diteliti. Fungsi diskriminan yang terbentuk digunakan untuk pengelompokan pada kasus baru yang belum diketahui keanggotaan kelompoknya. Nilai diskriminan dari setiap objek yang terlibat merupakan petunjuk termasuk dalam kelompok mana objek tersebut berasal (Johnson dan Wichern, 2007).

Pada umumnya metode yang sering digunakan untuk masalah klasifikasi pada penelitian sebelumnya adalah metode Diskriminan Linier yang dikembangkan oleh R. A. Fisher (1936). Metode diskriminan linier merupakan salah satu teknik multivariat yang berfokus pada pemisahan obyek dimana memerlukan asumsi variabel prediktor harus berdistribusi normal multivariat dan matriks varians kovarians harus sama. Tetapi pada penerapannya, metode diskriminan linier sering melibatkan variabel-variabel yang tidak mengikuti pola distribusi normal, sehingga diperoleh hasil klasifikasi diskriminan linier yang tidak optimal.

Untuk mengatasi hal tersebut maka dapat didekati dengan metode nonparametrik. Salah satu metode nonparametrik dalam hal klasifikasi adalah analisis diskriminan kernel yang menggunakan pendekatan fungsi kernel sehingga memungkinkan analisis diskriminan bekerja secara efisien dalam dimensi yang lebih tinggi (Mika *et al.*, 1999).

Terkait dengan peranan dan aplikasi analisis diskriminan yang merupakan salah satu dari teknik statistik multivariat yang banyak digunakan dalam berbagai bidang ilmu yang terjadi dalam sebuah fenomena sosial, kesehatan, keuangan, dan ekonomi. Hipertensi merupakan salah satu penyakit yang tergolong silent killer atau penyakit yang dapat membunuh manusia secara tidak terduga. Hipertensi sering disebut sebagai penyakit darah tinggi. Hal ini disebabkan, orang yang menderita hipertensi memiliki tekanan darah yang tinggi (abnormal) apabila diukur dengan menggunakan tensi meter. Melihat angka kunjungan hipertensi dengan jumlah yang cukup besar maka hipertensi perlu mendapat perhatian yang menyeluruh dan sungguh-sungguh. Hal inilah yang mendorong penulis untuk melakukan penelitian terhadap kejadian hipertensi dengan melakukan Analisis Diskriminan. Dimana Analisis Diskriminan berguna untuk mengklasifikasikan kelompok pasien yang hipertensi dan kelompok pasien tidak hipertensi dengan menggunakan faktor-faktor yang berhubungan dengan kejadian hipertensi sebagai acuan untuk menentukan kategori pasien.

2. Tinjauan Pustaka

2.1 Analisis Diskriminan

Analisis diskriminan adalah salah satu teknik statistik yang bisa digunakan pada hubungan dependensi (hubungan antar variabel dimana sudah bisa dibedakan mana variabel respon dan mana variabel prediktor).

2.1.1 Proses Dasar Analisis Diskriminan

Proses dasar analisis diskriminan:

1. Memisah variabel-variabel menjadi variabel dependen dan variabel independen.
2. Menentukan metode untuk membuat Fungsi Diskriminan, yakni:
 1. *Simultaneous estimation*, di mana semua variabel dimasukkan secara bersama-sama kemudian dilakukan proses Diskriminan.
 2. *Step-wise estimation*, di mana variabel dimasukkan satu persatu ke dalam model diskriminan. Pada proses ini, tentu ada variabel yang tetap ada pada model, dan ada kemungkinan satu atau lebih variabel independen yang 'dibuang' dari model.
3. Membentuk fungsi diskriminan
4. Menguji signifikansi dari Fungsi Diskriminan yang telah terbentuk, menggunakan *Wilk's Lambda* dan uji *F*.
5. Menguji ketepatan klasifikasi dari fungsi diskriminan.
6. Melakukan interpretasi terhadap Fungsi Diskriminan.
7. Melakukan uji validasi Fungsi Diskriminan.

2.1.2 Analisis Diskriminan Linear

Fungsi analisis diskriminan berkenaan dengan kombinasi linear yang bentuknya sebagai berikut:

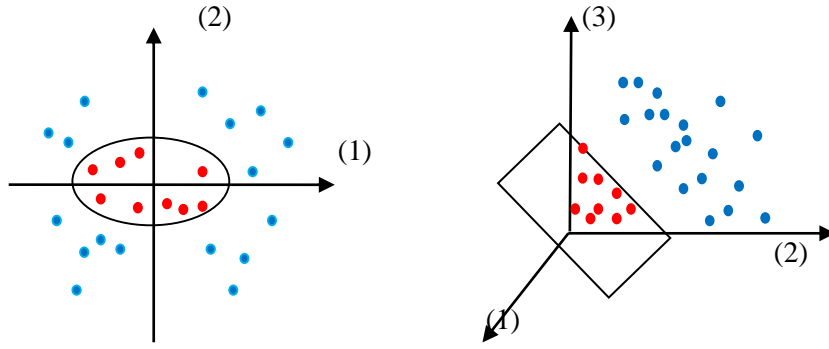
$$D_i = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \dots + b_jx_{ij} + \dots + b_px_{ip} \quad (2.1)$$

2.2 Metode Kernel

Metode Kernel merupakan suatu teknik pemetaan dataset dari ruang input (*input space*) ke ruang fitur (*feature space*) yang memiliki dimensi tinggi atau bahkan dimensi tak terbatas (*Hilbert space*) dengan pemetaan nonlinier ϕ . Pemetaan dari ruang input ke ruang fitur dinotasikan dengan :

$$\phi: x \rightarrow \phi(x).$$

Dan pemetaanya dapat dilihat pada gambar berikut.



Ruang Input X.

Ruang Fitur Dimensi Tinggi.

Gambar 2. 1 Fungsi ϕ memetakan data ke ruang vektor yang berdimensi lebih tinggi. Ilustrasi matriks kernel dapat dilihat pada **Gambar 2.2** berikut:

Dataset $S = \{x_1, \dots, x_l\}$

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & k(x_1, x_3) & \dots & k(x_1, x_l) \\ k(x_2, x_1) & k(x_2, x_2) & k(x_2, x_3) & \dots & k(x_2, x_l) \\ k(x_3, x_1) & k(x_3, x_2) & k(x_3, x_3) & \dots & k(x_3, x_l) \\ \dots & \dots & \dots & \dots & \dots \\ k(x_l, x_1) & k(x_l, x_2) & k(x_l, x_3) & \dots & k(x_l, x_l) \end{bmatrix} = k_{ik} = k(x_i, x_k),$$

dimana $i, k = 1, \dots, l$.

Gambar 2.2 Ilustrasi matriks kernel.

dinyatakan sebagai :

Suatu variabel random berdistribusi normal dengan mean μ dan variansi σ^2 jika

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), -\infty < x < \infty \quad (2.2)$$

Pada data multivariat, terlibat lebih dari satu variabel. Sekelompok variabel (X_1, X_2, \dots, X_p) dikatakan berdistribusi Normal p -variat dengan vektor mean $\mu = (\mu_1, \mu_2, \dots, \mu_p)$ dan matriks varians kovarians S jika fungsi joint distribusi dari p variabel ditentukan dengan rumus :

$$f(x) = \frac{1}{(2\pi)^{p/2} |S|^{1/2}} \exp\left(-\frac{1}{2} (x - \mu)' S^{-1} (x - \mu)\right), -\infty < x_i < \infty \quad (2.3)$$

Diketahui Fungsi Densitas Kernel pada kelompok t mengikuti distribusi normal adalah:

$$K_t(x) = \frac{1}{(2\pi)^{\frac{p}{2}} h^p |S_t|^{\frac{1}{2}}} \exp\left(-\frac{x^t S_t^{-1} x}{2h^2}\right) \quad (2.4)$$

Apabila $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$, maka estimator kernel dapat ditulis kembali sebagai:

$$\hat{f}_t(x) = \frac{1}{n_t} \sum_{i=1}^{n_t} K_t(x - x_{ti}) \quad (2.5)$$

Sedangkan estimator fungsi kepadatan kernel dinyatakan sebagai:

$$\hat{f}_t(x) = \frac{1}{n_t} \sum_{i=1}^{n_t} \frac{1}{(2\pi)^{\frac{p}{2}} h^p |S_t|^{\frac{1}{2}}} \exp\left(-\frac{(x-x_i)^t S_t^{-1} (x-x_i)}{2h^2}\right) \quad (2.6)$$

di mana:

n_t : banyak pengamatan pada kelompok ke- t

\mathbf{x} : nilai pengamatan yang akan diselidiki
 \mathbf{x}_i : nilai pengamatan pada kelompok ke- t
 S_t : matriks varians kovarians kelompok ke- t
 h : *bandwith*
 p : banyak variabel prediktor

Dalam hal ini h merupakan parameter pemulus (*smoothing parameter*) disebut juga dengan *bandwidth*. Nilai *bandwidth* antara 0,1 hingga 0,9. Masing-masing fungsi dibentuk dengan suatu nilai *bandwidth* yang optimal, agar mendapatkan hasil pendugaan yang baik. Nilai optimal h yang dihasilkan tergantung pada fungsi kepadatan dan kernel. Sehingga nilai *bandwidth* h optimal (Ansys, 2004):

$$h = \left(\frac{(A(k_t))}{n_t} \right)^{\frac{1}{p+4}}, \quad (2.7)$$

di mana konstanta optimal $A(k_t)$ tergantung pada kernel (k_t). Konstanta $A(k_t)$ dapat diperoleh dari:

Dengan kernel uniform

$$A(k_t) = \frac{4}{2p + 1} \quad (2.8)$$

Aturan klasifikasi pada Analisis Diskriminan Kernel menggunakan aturan *Bayes* berdasarkan probabilitas posterior terbesar (Johnson dan Wichern, 2007).

$$p(\pi_t | \mathbf{x}) = \frac{p_t \hat{f}_t(\mathbf{x})}{\sum_{t=1}^k p_t \hat{f}_t(\mathbf{x})} \quad (2.9)$$

2.4 Metode Cross Validation

Cross Validation merupakan salah satu teknik untuk memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Data yang digunakan dalam proses pembangunan model disebut data latih/*training*, sedangkan data yang akan digunakan untuk memvalidasi model disebut sebagai data *testing*. Jumlah data untuk *training* dan *testing* masing-masing adalah:

$$N_{training} = \frac{70}{100} \times N \quad (2.10 a)$$

$$N_{testing} = \frac{30}{100} \times N \quad (2.10 b)$$

Sedangkan untuk proporsi masing-masing kategori data keseluruhan adalah:

$$P_1 = \frac{1}{N} \times 100\% \quad (2.11 a)$$

$$P_2 = \frac{1}{N} \times 100\% \quad (2.11 b)$$

Proporsi masing-masing kategori dalam data *training* dan data *testing* harus sesuai dengan proporsi masing-masing kategori dalam data keseluruhan. Sehingga jumlah masing-masing kategori dalam data *training* dan *testing* adalah:

Untuk data *training*

$$N_1 = P_1 \times N_{training} \quad (2.12 a)$$

$$N_2 = P_2 \times N_{training} \quad (2.12 b)$$

Untuk data *testing*

$$N_1 = P_1 \times N_{testing} \quad (2.13 a)$$

$$N_2 = P_2 \times N_{testing} \quad (2.13 b)$$

2.5 Uji Ketepatan Klasifikasi

2.5.1 Apparent Error Rates (APER)

Tabel 2.2 Kelompok Prediksi

Kelompok Aktual	Kelompok Prediksi		Jumlah Observasi
	1	2	
1	n_{11}	n_{12}	n_1
2	n_{21}	n_{22}	n_2

Sumber: (Dian, 2015)

$$APER = \frac{n_{12} + n_{21}}{n_{11} + n_{22}} \quad (2.14)$$

2.6 Uji Keakuratan Pengelompokan dan Kestabilan

Proportional Change Criterion

$$C_{pro} = p^2 + q^2 \quad (2.15)$$

Maximum Change Criterion

$$C_{max} = (N_{max} / N) \times 100\% \quad (2.16)$$

Tahap pengujian selanjutnya yaitu menguji tingkat kestabilan pengelompokan dengan mengujikan pada *validation sample*. Dengan menghitung nilai *Press's Q* yang diformulasikan sebagai berikut (Hair *et al.*, 2006):

$$Press's Q = \frac{\{N - (nk)\}^2}{N(k - 1)} \quad (2.17)$$

3. Hasil dan Pembahasan

3.1 Data

3.1.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Puskesmas Usuku Kecamatan Tomia Timur Kabupaten Wakatobi Provinsi Sulawesi Tenggara Tahun 2013. Data tersebut terdiri dari 4 variabel prediktor yang terdiri dari faktor-faktor yang berhubungan dengan kejadian hipertensi (umur (X_1), obesitas (X_2), keteaturan olahraga (X_3), dan kebiasaan mengkonsumsi garam (X_4)), sedangkan variabel respon terdiri atas 2 variabel yang merupakan kelompok dengan kategori pasien yang hipertensi dan pasien yang tidak hipertensi.

3.2 Analisis Deskriptif

Tabel 3.1 Mean dan Standar Deviasi Data

Variabel	Tidak Hipertensi		Hipertensi	
	Mean	Standar Deviasi	Mean	Standar Deviasi
Umur	1,9062	0,2961	1,5147	0,50350
Obesitas	1,4688	0,5070	1,2647	0,4445
Keteraturan Olahraga	1,7188	0,4568	1,5882	0,4958
Konsumsi Garam	1,8750	0,3360	1,6471	0,4814

Tabel 3.1 menunjukkan bahwa pada kelompok tidak hipertensi variabel umur, obesitas, keteraturan olahraga, dan kebiasaan konsumsi garam mempunyai rata-rata yang lebih tinggi daripada kelompok hipertensi. Ini berarti pasien beresiko tinggi hipertensi apabila pasien tersebut mempunyai berat badan yang lebih dari IMT pasien ($IMT \geq 25$), mempunyai umur yang lebih dari 40 tahun, tidak teratur berolahraga yaitu berolahraga < 2 kali seminggu selama < 30 menit dalam sehari, dan mempunyai kebiasaan mengkonsumsi garam ≥ 6 sendok / minggu dibandingkan pasien yang tidak hipertensi.

3.3 Analisis Diskriminan

3.3.1 Uji Normal Multivariat

Tabel 3.2 Hasil Test Normality
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Umur	.412	100	.000	.607	100	.000
IMT	.428	100	.000	.593	100	.000
Keteraturan Olahraga	.407	100	.000	.611	100	.000
Konsumsi Garam	.453	100	.000	.562	100	.000

a. Lilliefors Significance Correction

Dengan menggunakan software SPSS 17 diperoleh hasil seperti pada **Tabel 3.2** semua variabel prediktor yaitu $0,00 < \alpha = 0,05$, maka data tidak berdistribusi normal. Karena pada prinsipnya pada analisis multivariat, variabel-variabel prediktornya dianggap suatu kesatuan dan tidak terpisah-pisah, jadi dapat disimpulkan bahwa data tidak berdistribusi normal multivariat.

3.3.2 Uji Homogenitas Varians Kovarians

Tabel 3.3 Hasil Test Box'M

Test Results	
Box's M	29.192
F Approx.	2.763
df1	10
df2	1.796E4
Sig.	.002

Tests null hypothesis of
equal population covariance
matrices.

Dengan menggunakan software SPSS 17 didapatkan hasil seperti pada **Tabel 3.3**. Karena nilai sig yaitu $0,002 < \alpha = 0,05$, jadi dapat disimpulkan bahwa matriks varians kovarians tidak homogen.

3.4 Metode Cross Validation

Dengan menggunakan metode *Cross Validation* maka dari 100 data pasien akan dibagi menjadi 2 partisi sebesar 70% untuk data training dan 30% untuk data *testing*. Sehingga jumlah data untuk data *training* dan testing masing-masing adalah:

$$N_{\text{training}} = \frac{70}{100} \times 100 = 70$$

$$N_{\text{testing}} = \frac{30}{100} \times 100 = 30$$

Jadi, untuk data *training* akan diambil secara acak 70 data pasien dan 30 data pasien untuk data *testing* dengan menggunakan Microsoft Excel dengan rumus: =RANDBETWEEN(1,100).

3.5 Analisis Diskriminan Kernel

Metode diskriminan kernel melakukan pengklasifikasian dari sebuah vektor pengamatan ke- \mathbf{x} didasarkan pada estimasi kepadatan dari kelompok. Metode kernel menggunakan *bandwidth* h dan fungsi kernel K_t untuk mengestimasi kepadatan kelompok t pada setiap vektor pengamatan ke- \mathbf{x} . Pada analisis diskriminan kernel ini digunakan fungsi kernel normal. Sedangkan dalam pemilihan *bandwidth*, akan dipilih berdasarkan

penghitungan *bandwith* optimal dimana konstanta optimal $A(K_t)$ tergantung pada Kernel (K_t) (Anslys, 2004).

3.5.1 Bandwith Optimal

Dengan menggunakan kernel normal ditentukan nilai $A(K_t)$ untuk data *training* dan *testing* dengan menggunakan persamaan (2.7) dan (2.8) masing-masing adalah:

$$A(K_t) = \frac{4}{2p+1} = \frac{4}{2 \cdot 4 + 1} = 0,444$$

Sehingga untuk h optimal diperoleh:

$$h = \left(\frac{(A(k_t))}{n_t} \right)^{\frac{1}{p+4}}$$

$$h_{training} = \left(\frac{0,444}{70} \right)^{\frac{1}{4+4}} = 0,531 \quad (3.1)$$

$$h_{testing} = \left(\frac{0,444}{30} \right)^{\frac{1}{4+4}} = 0,591 \quad (3.2)$$

3.5.2 Fungsi Diskriminan Kernel

$$\hat{f}_t(\mathbf{x}) = \frac{1}{n_t} \sum_{i=1}^{n_t} \frac{1}{(2\pi)^{\frac{p}{2}} h^p |\mathbf{S}_t|^{\frac{1}{2}}} \exp \left(-\frac{(\mathbf{x}-\mathbf{x}_i)^t \mathbf{S}_t^{-1} (\mathbf{x}-\mathbf{x}_i)}{2h^2} \right)$$

Untuk kelompok 1 :

$$\hat{f}_1(\mathbf{x}) = \frac{1}{18} \sum_{i=1}^{18} \frac{1}{(2\pi)^{\frac{p}{2}} h^p |\mathbf{S}_t|^{\frac{1}{2}}} \exp \left(-\frac{(\mathbf{x}-\mathbf{x}_i)^t \mathbf{S}_t^{-1} (\mathbf{x}-\mathbf{x}_i)}{2h^2} \right) \quad (3.3)$$

Untuk kelompok 2 :

$$\hat{f}_2(\mathbf{x}) = \frac{1}{52} \sum_{i=1}^{52} \frac{1}{(2\pi)^{\frac{p}{2}} h^p |\mathbf{S}_t|^{\frac{1}{2}}} \exp \left(-\frac{(\mathbf{x}-\mathbf{x}_i)^t \mathbf{S}_t^{-1} (\mathbf{x}-\mathbf{x}_i)}{2h^2} \right) \quad (3.4)$$

3.5.3 Probabilitas Posterior

Setelah diperoleh fungsi diskriminan kernel dari nilai penduga kepadatan kernel yang berasal dari kelompok 1 dan kelompok 2 pada, selanjutnya dilakukan perhitungan probabilitas posterior pengamatan \mathbf{x} dari kelompok 1 dan kelompok 2 dengan menggunakan persamaan (2.9).

Untuk Kelompok 1:

$$p(\pi_1|\mathbf{x}_1) = \frac{p_1 \hat{f}_1(\mathbf{x}_1)}{p_1 \hat{f}_1(\mathbf{x}_1) + p_2 \hat{f}_2(\mathbf{x}_1)}$$

Untuk Kelompok 2:

$$p(\pi_2|\mathbf{x}) = 1 - p$$

$$(\pi_1|\mathbf{x}) = \frac{p_2 \hat{f}_2(\mathbf{x})}{p_1 \hat{f}_1(\mathbf{x}) + p_2 \hat{f}_2(\mathbf{x})}$$

Tabel 3.4 Kelompok Prediksi Data *Training*

No	Kelompok Aktual	Kelompok Prediksi	No.	Kelompok Aktual	Kelompok Prediksi
1	1	1	36	1	1
2	2	2	37	2	2
3	2	2	38	2	2
4	2	2	39	1	1
5	2	2	40	2	2
6	2	2	41	2	2
7	2	2	42	2	2
8	2	2	43	2	2
9	2	2	44	1	1
10	2	2	45	2	2
11	1	1	46	2	2
12	2	2	47	2	2
13	2	2	48	1	1
14	2	2	49	1	1
15	2	2	50	2	2
16	1	1	51	2	2
17	2	2	52	2	2
18	2	2	53	2	2
19	2	2	54	1	1
20	2	2	55	1	1
21	1	1	56	2	2
22	2	2	57	2	2
23	1	1	58	2	2
24	1	1	59	1	1
25	1	1	60	2	2
26	2	2	61	1	1
27	2	2	62	2	2
28	2	2	63	2	2
29	2	2	64	2	2
30	2	2	65	2	2
31	2	2	66	2	2
32	1	1	67	2	2
33	2	2	68	2	2
34	2	2	69	2	2
35	1	1	70	2	2

Dari **Tabel 3.4** dapat dilihat bahwa tidak terdapat kesalahan klasifikasi.

3.5.4 Aplikasi pada Data *Testing*

Tabel 3.5 Kelompok Prediksi Data *Testing*

No	Kelompok Aktual	Kelompok Prediksi	No.	Kelompok Aktual	Kelompok Prediksi
1	2	2	16	2	2
2	2	2	17	1	1
3	2	2	18	1	1
4	2	1	19	2	1
5	2	2	20	1	1
6	2	2	21	2	2
7	2	2	22	1	1
8	2	2	23	1	1

9	2	2	24	1	1
10	1	1	25	2	2
11	1	1	26	1	1
29	1	1	27	1	1
13	2	2	28	2	2
14	2	2	29	1	2
15	1	2	30	1	1

Dari **Tabel 3.5** dapat dilihat pada kolom yang ditebalkan terdapat 2 pasien pada kelompok 2 yang salah klasifikasi, artinya 2 pasien yang diprediksikan masuk ke dalam kelompok 1 sedangkan pada data asli seharusnya masuk ke dalam kelompok 2. Kemudian terdapat 2 pasien pada kelompok 1 yang salah klasifikasi, artinya 2 orang pasien yang diprediksikan masuk ke dalam kelompok 2 sedangkan pada data asli seharusnya masuk ke dalam kelompok 1. Jadi total kesalahan klasifikasi adalah 4 unit percobaan.

3.6 Uji Ketepatan Klasifikasi

3.6.1 Peluang Kesalahan Klasifikasi (*APER*)

Untuk menguji berapa besar peluang kesalahan klasifikasi pada percobaan, digunakan persamaan (2.14) yang hasilnya dapat dilihat pada **Tabel 3.6** dan **Tabel 3.7**.

Untuk data *Training* :

Tabel 3.6 Hasil Klasifikasi data *Training*

Kelompok Aktual	Kelompok Prediksi		Jumlah Observasi
	1	2	
1	18	0	18
2	0	52	52

$$APER = \frac{0 + 0}{18 + 52} = 0$$

Artinya peluang kesalahan klasifikasi untuk data *Training* adalah sebesar 0.

Untuk data *Testing* :

Tabel 3.7 Hasil Klasifikasi data *Testing*

Kelompok Aktual	Kelompok Prediksi		Jumlah Observasi
	1	2	
1	12	2	14
2	2	14	16

$$APER = \frac{2 + 2}{14 + 16} = 0,133$$

Artinya peluang kesalahan klasifikasi untuk data *testing* adalah sebesar 0,133.

Sehingga total peluang kesalahan klasifikasi adalah *APER training* + *APER testing* yaitu $0 + 0,133 = 0,133$.

3.6.2 Uji Keakuratan

Hipotesis untuk menguji keakuratan klasifikasi yang dilakukan adalah:

H_0 : Klasifikasi akurat

H_1 : Klasifikasi tidak akurat

Statistik uji : Tolak H_0 apabila $C_{pro} < Hit_{ratio} < C_{max}$, artinya klasifikasi yang dilakukan belum akurat. Sebaliknya, terima H_0 apabila $Hit_{ratio} > C_{max} > C_{pro}$, artinya klasifikasi yang dilakukan sudah akurat.

Dengan menggunakan persamaan (2.15) dan (2.16) maka untuk data *Training*:

$$Hit_{ratio} = \frac{70}{70} \times 100 = 100\%$$

Proportional Change Criterion

$$C_{pro} = (0,743^2 + 0,257^2) 100\% = 61,796\%$$

Maximum Change Criterion

$$C_{max} = \left(\frac{52}{70}\right) \times 100\% = 74,3\%$$

Karena nilai $Hit_{ratio} > C_{max} > C_{pro}$ yaitu $100\% > 74,3\% > 61,796\%$ maka H_0 diterima, yang artinya hasil klasifikasi akurat.

Training:

Untuk data *Testing*:

$$Hit_{ratio} = \frac{26}{30} \times 100\% = 86,67\%$$

Proportional Change Criterion

$$C_{pro} = (0,533^2 + 0,467^2) 100\% = 50,222\%$$

Maximum Change Criterion

$$C_{max} = \left(\frac{16}{30}\right) \times 100\% = 53,3\%$$

Karena nilai $Hit_{ratio} > C_{max} > C_{pro}$ yaitu $86,667\% > 53,3\% > 50,222\%$ maka H_0 diterima, yang artinya hasil klasifikasi akurat.

3.6.3 Uji Kestabilan

Hipotesis untuk menguji kestabilan klasifikasi yang dilakukan adalah:

H_0 : Klasifikasi konsisten

H_1 : Klasifikasi tidak konsisten

Statistik uji : Tolak H_0 apabila nilai $Press's Q < \chi^2_{(db,\alpha)}$ artinya klasifikasi yang dilakukan belum konsisten. Sebaliknya, terima H_0 apabila $Press's Q < \chi^2_{(db,\alpha)}$ artinya klasifikasi yang dilakukan sudah konsisten.

Dengan menggunakan persamaan (2.15) maka diperoleh:

$$Press's Q = \frac{[100 - (88 \times 2)]^2}{100(2 - 1)} = 84,64$$

Karena nilai statistik uji $Press's Q < \chi^2_{(1,0,05)}$ yaitu $84,64 > 3,841$ dapat disimpulkan bahwa keakuratan pengklasifikasian adalah konsisten.

4. Kesimpulan

4.1 Kesimpulan

Berdasarkan uraian dan pembahasan dalam penelitian ini, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Analisis Diskriminan Kernel dengan metode *Cross Validation* dilakukan dengan membagi data menjadi dua bagian yaitu sebagian data digunakan untuk *training* yang bertujuan membentuk model dan sebagian lainnya digunakan untuk *testing* yang bertujuan untuk validasi model.
2. Dari hasil klasifikasi dengan menggunakan metode kernel diketahui bahwa pada pasien yang hipertensi adalah 68 dan pasien yang tidak hipertensi 32.
1. Dari hasil analisis diskriminan kernel diperoleh nilai $Press's Q < \chi^2_{(1,0,05)}$ yaitu $84,64 > 3,841$ dapat disimpulkan bahwa pengklasifikasian dengan diskriminan kernel mempunyai kestabilan dalam membedakan dan memprediksi pasien yang tidak hipertensi dan yang hipertensi.

DAFTAR PUSTAKA

- Ansys, Inc., 2004. *Ansys Theory Reference : Ansys Release 9.0, First Ed.* SAS IP, Inc, Philadelphia.
- Arisona, Dian Christien. 2015. *Analisis Diskriminan Linier pada Klasifikasi Nasabah Menunggak dan Tidak Menunggak dengan Metode Cross Validation.* Makassar : Program Sarjana, Universitas Hasanuddin.

- Fitria , Yosiana. W, Bambang Widjanarko Otok. 2011. *Evaluasi Ketepatan Klasifikasi Kelulusan Tes Keterampilan SNMPTN Bidang Olahraga Menggunakan Analisis Diskriminan Kernel*. Surabaya: Program Sarjana. Universitas Teknologi Surabaya.
- Hair J. F., R. E. Anderson, R. L. Tatham, W. C. Black. 2006. *Multivariate Data Analysis*. Sixth Edition, Pearson Education Prentice Hall, Inc.
- Johnson, N. and Wichern, D. 2007. *Applied Multivariate Statistical Analysis, 6th Edition*. New Jersey: Prentice Hall, Englewood Cliffs.
- Jones, M. C. and Wand, M. P. 1995. *Kernel Smoothing*. Chapman and Hall.
- Juliati. 2013. *Faktor-Faktor yang Berhubungan dengan Kejadian Hipertensi Diwilayah Kerja Puskesmas Usuku Kecamatan Tomia Timur Kabupaten Wakatobi Provinsi Sulawesi Tenggara Tahun 2013*. Makassar: Program Sarjana. Universitas Muslim Indonesia.
- Meilianawati, Putri, Eni Sumarminingsih, dan Ni Wayan Surya Wardhani. 2012. *Pendekatan Model Proportional Odds dan Analisis Diskriminan Kernel pada Regresi Respon Ordinal*. Malang: Program Sarjana, Universitas Brahwijaya.
- Mika. S, G. Rˆatsch, J. Weston, B. Schˆolkopf, and K.-R. Mˆuller. 1999 . Fisher discriminant analysis with kernels. In Y.-H. Hu, J.Larsen, E. Wilson, and S. Douglas, editors, *Neural Networks for Signal Processing IX*, pages 41–48. IEEE.
- Suparti dan Sudargo. 2006. *Estimasi Densitas Mulus dengan Metode Kernel*. Semarang: Program Sarjana, Universitas Diponegoro.